

УДК 372.881.1

УДК 004

Н.А. СИГАЧЕВА,

*кандидат педагогических наук, доцент
Казанский федеральный университет*

М. КИКУ,

*студент
Казанский федеральный университет*

Р.А. КЛИМОВ,

*студент
Казанский федеральный университет*

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ПЕРЕВОДЕ ПРОФЕССИОНАЛЬНО- ОРИЕНТИРОВАННЫХ, АУТЕНТИЧНЫХ ТЕКСТОВ

Аннотация. В статье рассматриваются проблемы использования информационных технологий в переводе аутентичных профессиональных технических текстов. Целью работы является выбор оптимальной системы машинного перевода технической литературы. Авторы анализируют различные подходы к осуществлению машинного перевода, сравнивают и предлагают пути решения данной проблемы. Результатом данного исследования является аргументированный выбор программ для машинного перевода технических текстов.

Ключевые слова: информационные технологии; перевод; машинный перевод; статистический машинный перевод (СМТ); перевод «по правилам»; аутентичный текст; профессионально-ориентированный текст.

Abstract. The article deals with the problems of using information technologies in the translation of authentic professional technical texts. The aim of the work is to select the optimal machine translation system for technical literature. The authors analyze different approaches to the implementation of machine translation, compare and suggest ways to solve this problem. The result of this study is a reasoned choice of programs for machine translation of technical texts.

Keywords: Information technology, translation; machine translation; statistical machine translation (SMT); “rule-based” machine translation (RBMT); authentic text; professionally-oriented text.

В настоящее время проблема использования информационных технологий в переводе является актуальной, поскольку быстрое раз-

витие науки вызывает появление новых технических средств, документация которых, как правило, пишется на английском языке. Перевод с иностранного языка без использования информационных технологий требует больших временных и денежных ресурсов.

Технология машинного перевода непрерывно развивается и сегодня может предложить более быстрый и дешевый способ решения данной проблемы. Разработки технологий начались в 50-х годах прошлого века и до настоящего времени, успешно прошли ряд этапов в своем развитии. До 70-х годов, из-за низкой производительности ЭВМ того времени, было сложно, практически неосуществимо выполнять рентабельный качественный перевод. Одновременно, развитие вычислительной техники содействовало продолжению исследований. Важным аспектом явились разработанные персональные компьютеры, сложные словарные и поисковые системы, обрабатывающие языковые данные. Увеличилась необходимость в переводах в соответствии с возрастанием международных взаимосвязей. Вышеуказанное привело к активизации этой области научных исследований, с середины 1970-х. В 1980-е годы для переводческих систем наступил эра широкого применения на практике, стало возможным пользоваться коммерческими разработками [2].

Принято, процесс использования программного обеспечения для осуществления переводческой деятельности, то есть переноса информации с исходного языка, на другой, без участия человека называть машинным переводом.

Несомненно, сейчас существует и разрабатывается большое количество проектов машинного перевода на коммерческой основе. Некоторые крупные компании, такие как Google и Yandex предоставляют возможность пользоваться последними достижениями в данной сфере бесплатно. В России исследовательская группа профессора Р. Г. Пиотровского осуществила существенный вклад в продвижение систем машинного перевода [2].

Рассмотрим два основных типа машинного перевода: Перевод «по правилам» и Статистический машинный перевод [1].

Первый подход называют классическим или традиционным. Его применяют многие дизайнеры систем машинного перевода (например, российская компания PROMT). Это термин, который обозначающий системы машинного перевода на базе языковых данных о двух языках. Речь идет о включении двуязычных словарей и грамматик, охватывающих наиболее важные семантические, морфологические, синтак-

сические закономерности каждого из языков. С использованием вышеуказанных данных возможно преобразовывать исходный текст, в его перевод [3].

Исследователи определяют статистический машинный перевод как вид машинного перевода текста, основанного на сравнительном анализе больших количеств языковых пар [5]. SMT в автоматическом режиме адаптируется к новой ситуации в переводе. Системы обращают внимание на перевод новых слов, отдельных выражений, в том числе, если их перевод изменялся по сравнению с предыдущим. Данным ресурсам необходим корпус для качественной работы. Системы без затруднений «запоминают» новые варианты перевода и используют их в последующей работе. Программа как бы «живет» с развитием перевода. Процесс увеличивает скорость ее «обучения» и способствует повышению качества окончательного перевода. Дизайнеры систем подчеркивают, что наличие большого количества языковых пар ресурса и более точное взаимосоответствие, предлагает лучший вариант статистического машинного перевода [5].

Ко второму типу машинных переводчиков относится Google Translate. При переводах Google применяются алгоритмы, основанные на теории вероятности и статистике. При этом программа обладает значительным и постоянно растущим словарным запасом, насчитывающим на настоящий момент более триллиона слов [4].

Рассмотрим два вида систем на примере. SMT требует анализировать параллельные корпуса для осуществления перевода. RBMT нуждается в «знаниях», которые вкладывают в нее программисты: грамматику, синтаксис, семантику, лексику.

Для примера используем предложение из документации Python 3.6.3. В качестве SMT системы воспользуемся сервисом Google Translate, в качестве RBMT – переводчиком от PROMT, предназначенным для перевода технической литературы.

Попытаемся осуществить перевод данного предложения: «The break statement, like in C, breaks out of the inner most enclosing for or while loop.» Система SMT предлагает следующий перевод: «Оператор break, как и на C, вырывается из самого внутреннего окружения для цикла или while».

Другой вариант дает система RBMT: «Оператор break, как и на C, вырывается из самого внутреннего окружения цикла for или while».

Пример показывает, что SMT система перевела слово «for», несмотря на то, что в данном примере это не следовало делать. RBMT

система, созданная специально для перевода текстов из области информационных технологий, перевела все правильно. При дальнейшем сравнительном анализе двух систем выявляется целый ряд их преимуществ и недостатков (см. таблицу 1).

Системы типа RBMT	Системы типа SMT
+ Наличие определенной предсказуемости в результатах переводов	- Невозможно предсказать конечные результаты переводов
+ Качественный перевод текстов общей тематики	+ Качественный перевод текстов специальной тематики
+ Наличие «Знаний» о грамматических правилах	- Работа «не по правилам»
+ Не предъявляются требования к аппаратному обеспечению	- Необходимое сильное аппаратное обеспечение
- Излишек буквализма	+ «Живой» перевод
- Необходимы большие вложения разработчиков, а также специальные знания простого пользователя	+ Благодаря достаточно объемному корпусу возможно ускорение и улучшение работы программы, без дополнительных действий
+ Можно изменять исходный вариант с целью повышения качества перевода	- Строгая «привязка» к словарному запасу и невозможность обнаружить в нем соответствия не дают возможности осуществлять изменения и совершенствовать качество переводимого текста

Табл. 1. Преимущества и недостатки RBMT и SMT подходов при переводе с иностранного языка

Подводя итоги, можно отметить, что для перевода технических текстов оптимальным является RBMT подход. В документациях используется большое количество терминов, а предложения состоят из уникальных словосочетаний. Для SMT подхода это практически невыполнимо, так как невозможно составить конструкции, удовлетворяющие всем технологическим направлениям, существующим сегодня.

Литература

1. Филиппович А.Ю. Системы автоматического (машинного) перевода текста. – URL: <http://it-claim.ru/Education/Course/Lingvistika/Lecture/Lecture13.pdf> (дата обращения: 27.09.2017).
2. Машинный перевод. – URL: https://en.wikipedia.org/wiki/Machine_translation (дата обращения: 27.09.2017).

3. Машинный перевод на основе правил.– URL: https://en.wikipedia.org/wiki/Rule-based_machine_translation (дата обращения: 27.09.2017).
4. Статистический машинный перевод. – URL: https://en.wikipedia.org/wiki/Statistical_machine_translation (дата обращения: 27.09.2017).
5. Исламов Р.С., Фомин А.Г. Анализ современных систем машинного перевода типа SMT И RBMT. – URL: http://scjournal.ru/articles/issn_1997-2911_2013_3-1_18.pdf (дата обращения: 27.09.2017)